ENGINEERING RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN
ANN ARBOR

THE THEORY OF SIGNAL DETECTABILITY

PART I. THE GENERAL THEORY

ISSUED SEPARATELY:

PART II APPLICATIONS WITH GAUSSIAN NOISE

Technical Report No. 13
Electronic Defense Group
Department of Electrical Engineering

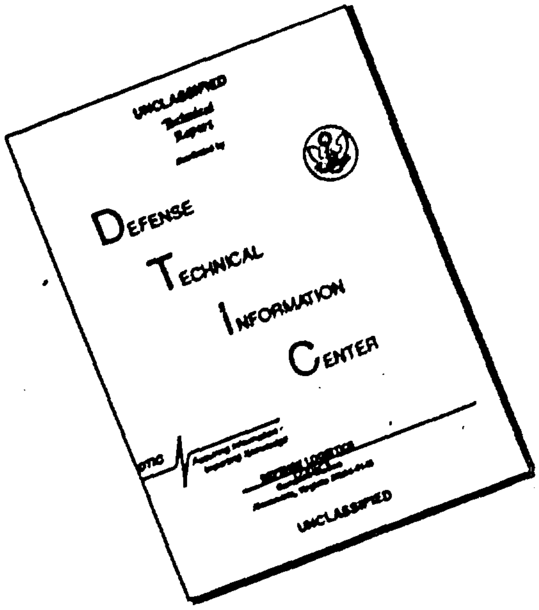
By: W. W. Peterson
T. G. Birdsall

Approved by H. W. Welch, Jr.

M970, TASK NO. EDG-3 CONTRACT NO. DA-36-0...
...RTMENT OF THE ARMY, DEPARTMENT OF ARMY 039 sc-15358, SIGNAL
... SIGNAL CORPS PROJECT NO. 29-194B-0. ... PROJECT NO.

JUNE, 1...

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

ERRATA . . . . TECHNICAL REPORT NO. 13

Part I

pg. 10. The notation $A_2(k)$ mean . . . . a criterion such that $P_N(A_2(k)) = k$. This notation is defined on page 16.

pg. 1?, line 3, The sentence should re . . . at any point $(P_N(A), P_{SN}(A))$ on curve (1) a line is drawn with slope β_k giv . . . the operating level graph, it will be tangent to the curve and will intersec . . . axis at the value $P_{SN}(A) - \beta_k P_N(A)$."

pg. 24, the second line from the bot . . . of the page should start "$P_{SN}(A_2 - A_1) = 0$"

pg. 35, Omit the x_0 between lines 4 . . .

pg. 40, Line 6 should read "measurab . . . contained in A_0 such that $P(B_0) = \gamma$."

Part II

pg. 3, Footnote 2 should read "If $\frac{1}{\sqrt{2\pi}}$ . . . . etc."

pg. 37, line 3 should read "times the . . . tude squared of its envelope, etc."

pg. 64 line 1, replace "when" by "For . . .

Note: An introduction to the theory . . . gnal detectability using as little mathematics as possible and including discussions of the applications of sequential analysis as well as the types of . . . criteria discussed in Part I is being prepared as EDG Technical Report . . . 4. Enough theoretical material will be included so that this report could be . . . in place of Part I as an introduction to Part II.

ENGINEERING RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN
ANN ARBOR

THE THEORY OF SIGNAL DETECTABILITY

PART I. THE GENERAL THEORY

ISSUED SEPARATELY:

PART II. APPLICATIONS WITH GAUSSIAN NOISE

Technical Report No. 13
Electronic Defense Group
Department of Electrical Engineering

By: W. W. Peterson
T. G. Birdsall

Approved by: H. W. Welch
H. W. Welch, Jr.

Project M970

TASK ORDER NO. EDG-3
CONTRACT NO. DA-36-039 sc-15358
SIGNAL CORPS, DEPARTMENT OF THE ARMY
DEPARTMENT OF ARMY PROJECT NO. 3-99-04-042
SIGNAL CORPS PROJECT NO. 29-194B-0

June, 1953

TABLE OF CONTENTS

PART I. THE GENERAL THEORY

TABLE OF CONTENTS (cont.)

PART II. APPLICATIONS WITH GAUSSIAN NOISE (Issued Separately)

ABSTRACT

PART I

The several statistical approaches to the problem of signal detectability which have appeared in the literature are shown to be essentially equivalent. A general theory based on likelihood ratio embraces the criterion approach, for either restricted false alarm probability or minimum weighted error type optimum, and the a posteriori probability approach. Receiver reliability is shown to be a function of the distribution functions of likelihood ratio. The existance and uniqueness of solutions for the various approaches is proved under general hypothesis.

PART II

The full power of the theory of signal detectability can be applied to detection in Gaussian noise, and several general results are given. Six special cases are considered, and the expressions for likelihood ratio are derived. The resulting optimum receivers are evaluated by the distribution functions of the likelihood ratio. In two of the special cases studied, the uncertainty of the signal ensemble can be varied, throwing some light on the effect of uncertainty on probability of detection.

ACKNOWLEDGEMENTS

In the work reported here, the authors have been influenced greatly by their association with the other members of the Electronic Defense Group. In particular, Mr. H. W. Batten contributed much to the early phases of the work on signal detectability. The authors are indebted to Mr. W. C. Fox and Mr. Paul Roth for the proofs of Lemma 1 and Lemma 2 in Appendix B and also to Mr. Fox for the proof of Lemma 4 and for the many helpful suggestions and corrections resulting from his careful reading of the text.

The authors also wish to acknowledge their indebtedness to Dr. A. B. Macnee, Dr. H. W. Welch, and Mr. C. B. Sharpe for the many suggestions resulting from their reading the report, and to Geraldine L. Preston for her assistance in the preparation of the text.

THE THEORY OF SIGNAL DETECTABILITY

PART I. THE GENERAL THEORY

ISSUED SEPARATELY:

PART II. APPLICATIONS WITH GAUSSIAN NOISE

1. Concepts and Theoretical Results

1.1 Introduction

Random interference plays the key role in the theory of signal detectability. It not only places a limit on the energy which a signal must have to be detected reliably, but it also limits the bandwidth of a receiver for strong signals, or generally the variety of signals which can be detected consistently in a given receiver. Part I of this report presents the basic theory of detecting signals in random interference and Part II applies it to some simple problems in design and evaluation of receivers.

The signal detectability problem is represented schematically in Fig. 1.1. The operator has available a voltage varying with time, which will be referred to as the receiver input. This voltage is in some way different when a signal is present from when there is noise alone.

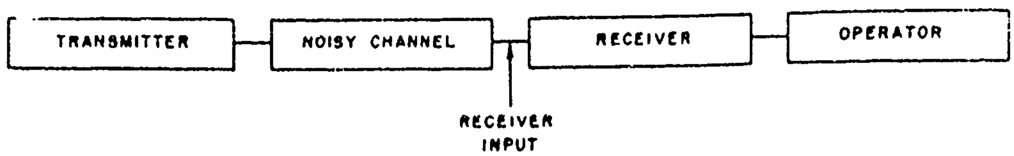


FIG. 1.1. BLOCK DIAGRAM OF SIGNAL DETECTION PROBLEM.

The receiver is the operator's tool or analyzing system, it enables him to study the input to the receiver by observing the receiver output. He can use the receiver input to his advantage only if (1) the receiver input is different when there is a signal than when there is no signal, and (2) he knows enough about the signals and the noise to analyze the input so as to recognize the difference. The operator can do better than random guessing in deciding whether or not there is a signal present only when he has information about the signals, the noise, and his receiver; this must be recognized before treating this problem. The information about the signal and about the noise is usually of a statistical nature because of the random nature of noise, and the uncertainty as to the exact signal that will be transmitted.

Signal detectability has been recognized as a statistical problem by a number of authors.[1] There have been two distinct approaches to the problem. The first, the criterion approach, is first presented in <u>Threshold Signals</u> by J. L. Lawson and G. I. Uhlenbeck.[2] The second, using a posteriori probability,

---

[1] Lawson and Uhlenbeck, Ref. 1; Woodward and Davies, Refs. 2, 3, 4, and 5; Reich and Swerling, Ref. 6; Middleton, Ref. 7; Slattery, Ref. 8; Hanse, Ref. 9; Schwartz, Ref. 10; North, Ref. 11; Kaplan and Fall, Ref. 12.

[2] Lawson and Uhlenbeck, Ref. 1.

was studied by P. M. Woodward and I. L. Davies.[1] The difference between the two methods lies mainly in the approach. Both are presented in this report, and the very close connection between the results of the two will be demonstrated in Section 2; namely, the basic receiver required can be the same for either case, only the final manner of analysis and presentation of the output is different. The criterion approach requires less of this analysis, and has been given more attention in this report because it is somewhat simpler.

1.2 Detectability Criteria

Suppose the operator is required to guess whether or not there is a signal present. He will, for certain receiver inputs, say that a signal is present.[2] Such receiver inputs will be said to satisfy the criterion, or to be in the criterion. Those receiver inputs which lead him to guess that there is no signal present are not in the criterion.

There are two distinct kinds of errors which the operator may make. He may say there is a signal present if there is only noise; this is a false alarm. He may say there is only noise when signal plus noise is present; he misses the signal. One of these errors may be more serious than the other, so that they must be considered separately.

It will be convenient to use the ordinary notation of probability theory. Events will be represented by letters, and in particular, the following symbols will be used for the following events:

---

[1]Davies, Ref. 2., and Woodward and Davies, Ref. 3.

[2]We shall assume the operator is scientifically logical, i.e., for the same receiver input he will always give the same response. An alternative approach is described in Appendix A.

SN There is signal plus noise
N There is noise alone
A The operator says there is a signal, i.e., the receiver input is in the criterion
CA The operator says there is only noise, i.e., the receiver input is not in the criterion.

If B and C are events, P(B) is the probability of occurrence of event B, P(B·C) is the probability of occurrence of events B and C together, and $P_B(C)$ is the (conditional) probability of occurrence of event C if event B is known to occur.

From the statistical information given about the signal and the interference it turns out to be convenient to calculate $P_N(A)$ and $P_{SN}(A)$, because these quantities do not depend upon the a priori probability that a signal is present. This will be done in Part II of this report for some interesting cases. If these probabilities, $P_N(A)$ and $P_{SN}(A)$ are given as well as P(SN), the a priori probability that a signal is present, then the probability of any combination of the events in this discussion can be calculated. In fact, any three (algebraically) independent probabilities can be used to calculate all the others. That there are just three (algebraically) independent probabilities can be seen by noting that all of the events discussed are combinations of the four events SN·A, N·A, SN·CA, and N·CA, and any probabilities can be calculated from the probabilities of these four. But the sum of the probabilities of these four is unity, so only three of these are independent. Thus, for example,

$$
\begin{aligned}
P(SN\cdot A) &= P(SN)\, P_{SN}(A)\ , \\
P(N\cdot A) &= \left[1 - P(SN)\right] P_N(A), \\
P(SN\cdot CA) &= P(SN)\, P_{SN}(CA) = P(SN)\left[1 - P_{SN}(A)\right]\ , \qquad (1.1) \\
P(A) &= P(SN\cdot A) + P(N\cdot A), \\
P_A(SN) &= \frac{P(SN\cdot A)}{P(A)}\ , \text{etc.}
\end{aligned}
$$

1.2 A Posteriori Probability and Signal Detectability

As an alternative to requiring the operator to say whether a signal is present or not, the operator might be asked what, to the best of his knowledge, is the probability that a signal is present. This approach has the advantage of getting more information from the receiving equipment. In fact Woodward and Davies point out that if the operator makes the best possible estimate of this probability for each possible transmitted message, he is supplying all the information which his equipment can give him.[1] The method of making the best estimate of the a posteriori probability that a signal is present will be discussed in this report. A good discussion of this approach is also found in the original papers by Woodward and Davies.[2]

It is shown in Section 2 that the a posteriori probability is given by the following equation:

$$P_x(SN) = \frac{\ell(x)\ P(SN)}{\ell(x)\ P(SN) + 1 - P(SN)} \tag{1.2}$$

where $P_x(SN)$ is the a posteriori probability for the receiver input denoted by x and $\ell(x)$ is the likelihood ratio for the same receiver input. Likelihood ratio for a particular receiver input is usually defined as the ratio of probability density for that receiver input if there is signal plus noise to the probability density if there is noise alone. It is a measure of how likely that receiver input is when there is signal plus noise as compared with when there is noise alone. It is a random variable; its value depends upon what the receiver input happens to be. If a receiver which has likelihood ratio as its output

[1] Ref. 3.

[2] Ref. 2, 3, 4, and 5.

can be built, and if the a priori probability $P(SN)$ is known, a posteriori probability can be calculated easily. The calculation could be built into the receiver calibration, making the receiver an optimum receiver for obtaining a posteriori probability.

1.4 Optimum Criteria

An important question is whether or not it is possible to find the optimum criterion for a given situation. A first step toward the answer is to define what is meant by optimum, and this definition depends upon the situation. It may be possible to put a numerical value upon the correct responses and a numerical cost on the errors. Suppose

$$\begin{aligned} V_{SN \cdot A} &= \text{Value of the correct response } SN \cdot A \\ V_{N \cdot CA} &= \text{Value of the correct response } N \cdot CA \\ K_{SN \cdot CA} &= \text{Cost of the error } SN \cdot CA \\ K_{N \cdot A} &= \text{Cost of the error } N \cdot A \end{aligned} \tag{1.3}$$

Then

$$V = V_{SN \cdot A} P(SN \cdot A) + V_{N \cdot CA} P(N \cdot CA) - K_{SN \cdot CA} P(SN \cdot CA) - K_{N \cdot A} P(N \cdot A) \tag{1.4}$$

is the expected value of the response of the equipment for a given criterion. An optimum criterion then would be one which would maximize this expression. Since the later sections will calculate $P_N(A)$ and $P_{SN}(A)$, it will be an advantage to express the expected value V of the response in terms of these quantities.

$$V = V_{SN \cdot A} P(SN)\, P_{SN}(A) + V_{N \cdot CA}\left[1 - P(SN)\right]\left[1 - P_N(A)\right] - K_{SN \cdot CA} P(SN)\left[1 - P_{SN}(A)\right] - K_{N \cdot A}\left[1 - P(SN)\right] P_N(A)$$

$$V = P_{SN}(A)\, P(SN)\, (V_{SN \cdot A} + K_{SN \cdot CA}) - P_N(A)\left[1 - P(SN)\right](V_{N \cdot CA} + K_{N \cdot A}) + V_{N \cdot CA}\left[1 - P(SN)\right] - K_{SN \cdot CA}\, P(SN). \quad (1.$$

Thus maximizing V is equivalent to requiring that

$$P_{SN}(A) - \beta\, P_N(A) \text{ is a maximum, where}$$

$$\beta = \frac{1 - P(SN)}{P(SN)} \; \frac{(V_{N\cdot CA} + K_{N\cdot A})}{(V_{SN\cdot A} + K_{SN\cdot CA})} \quad . \tag{1.6}$$

Note that P(SN) is the a priori probability that there is a signal present.

In another case it may be required to limit the probability of a false alarm and to minimize the probability of a missed signal with this restriction. In symbols, it is required that,

$$\begin{aligned} &P(N\cdot A) \leq Po \\ &P(SN\cdot CA) \text{ is a minimum.} \end{aligned} \tag{1.7}$$

This also can be expressed in terms of $P_N(A)$, $P_{SN}(A)$, and the a priori probability P(SN):

$$\begin{aligned} &P(N\cdot A) = \left[1 - P(SN)\right] P_N(A) \leq Po, \quad \text{or} \quad P_N(A) \leq k = \frac{Po}{1 - P(SN)} \;, \text{ and} \\ &P(SN\cdot CA) = P(SN)\left[1 - P_{SN}(A)\right] \text{ is a minimum, i.e., } P_{SN}(A) \text{ is a maximum.} \end{aligned} \tag{1.8}$$

1.5 Theoretical Results

Both of the above problems of finding an optimum criterion will be discussed in later sections, and it will be shown that under very general conditions both problems have essentially the same solution. The optimum criterion consists of all receiver inputs with likelihood greater than some number β. For the first type of optimum criterion, β is the parameter in Eq. (1.6), and for the second type of criterion, β can be determined from the value of the parameter k in Eq. (1.8). It has already been mentioned that a posteriori probability is the simple function of likelihood ratio given in Eq. (1.2). Thus a receiver which could calculate the likelihood ratio for each receiver input can be used as an a posteriori probability type receiver or as either of the criterion type receivers. Part II of this report, which treats some specific cases, deals only with the likelihood ratio.

1.6 Receiver Evaluation[1]

Usually a receiver is judged on the basis of probability of false alarm if no signal is sent, i.e., $P_N(A)$, and the probability of detection if a signal is sent, $P_{SN}(A)$. The reliability of any receiver in any given situation can be summarized in one graph, called the receiver operating characteristic, on which $P_{SN}(A)$ is plotted against $P_N(A)$. For any criterion and any fixed set of signals, there is fixed value for $P_{SN}(A)$ and a fixed value for $P_N(A)$. Thus the criterion can be represented as a point on the receiver operating characteristic graph. A criterion-type receiver may operate at any level (i.e., any value of β or any value of K), and hence is represented by a curve. Two types of optimum criteria have been discussed, and the graph points up the relation between the two. In Fig. 1.2 curve (1) is based on optimum operation for which $P_{SN}(A)$ is maximized for $P_N(A)$ fixed. Thus, no receiver can operate above the first curve. The third curve is a lower limit in operation found by rotating the optimum curve about the center point of the graph; it would result if an optimum receiver operator minimized $P_{SN}(A)$, i.e., said no whenever he should say yes, and vice versa. No receiver, no matter how poor, can be made to operate below the third curve. The diagonal could be achieved by turning the receiver off and guessing, in which case $P_{SN}(A) = P_N(A)$.

In the next section it will be shown that the derivative of curve (1) sketched in the lower plot, is the operating level β of the optimum receiver; that is, if the slope at some point is β, then the corresponding optimum criterion

[1] Only evaluation of criterion type receivers is discussed here. Evaluation of an a posteriori probability type receiver is considered in Section 2.5.

M-970 A-63-67 JEM 6-11-53

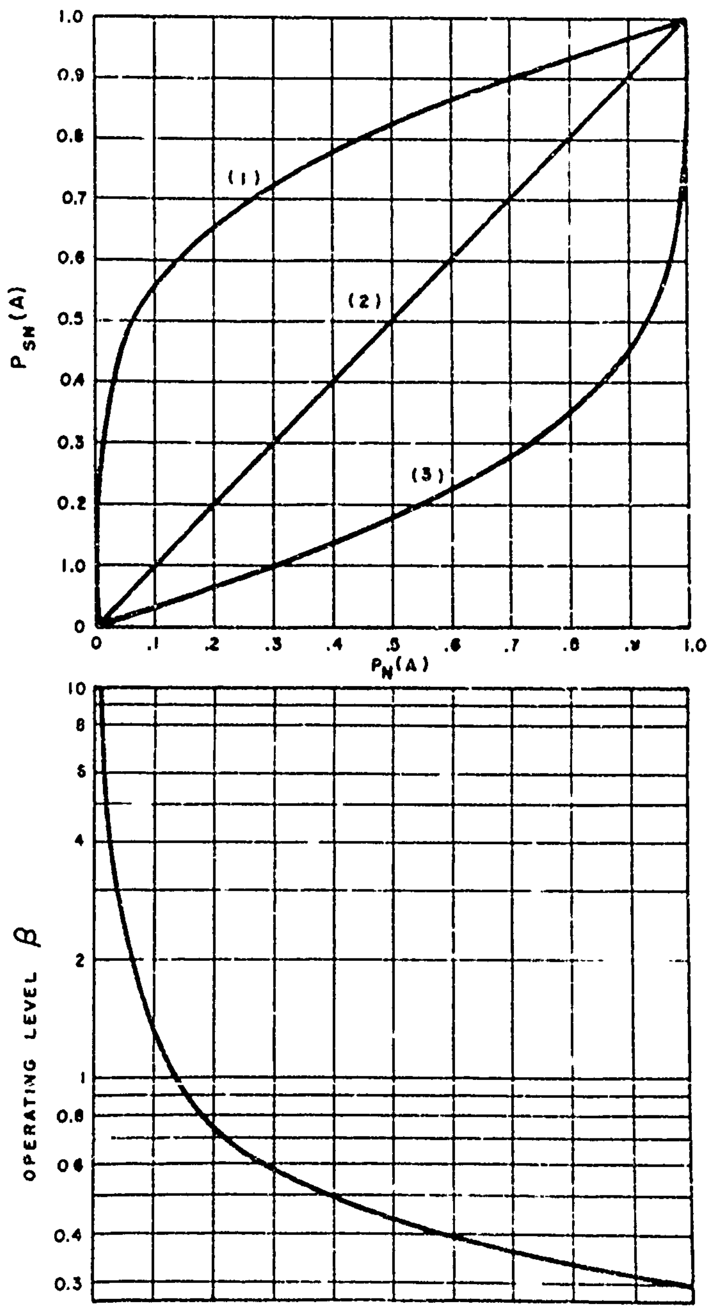


FIG. 1.2

TYPICAL RECEIVER OPERATING CHARACTERISTIC.

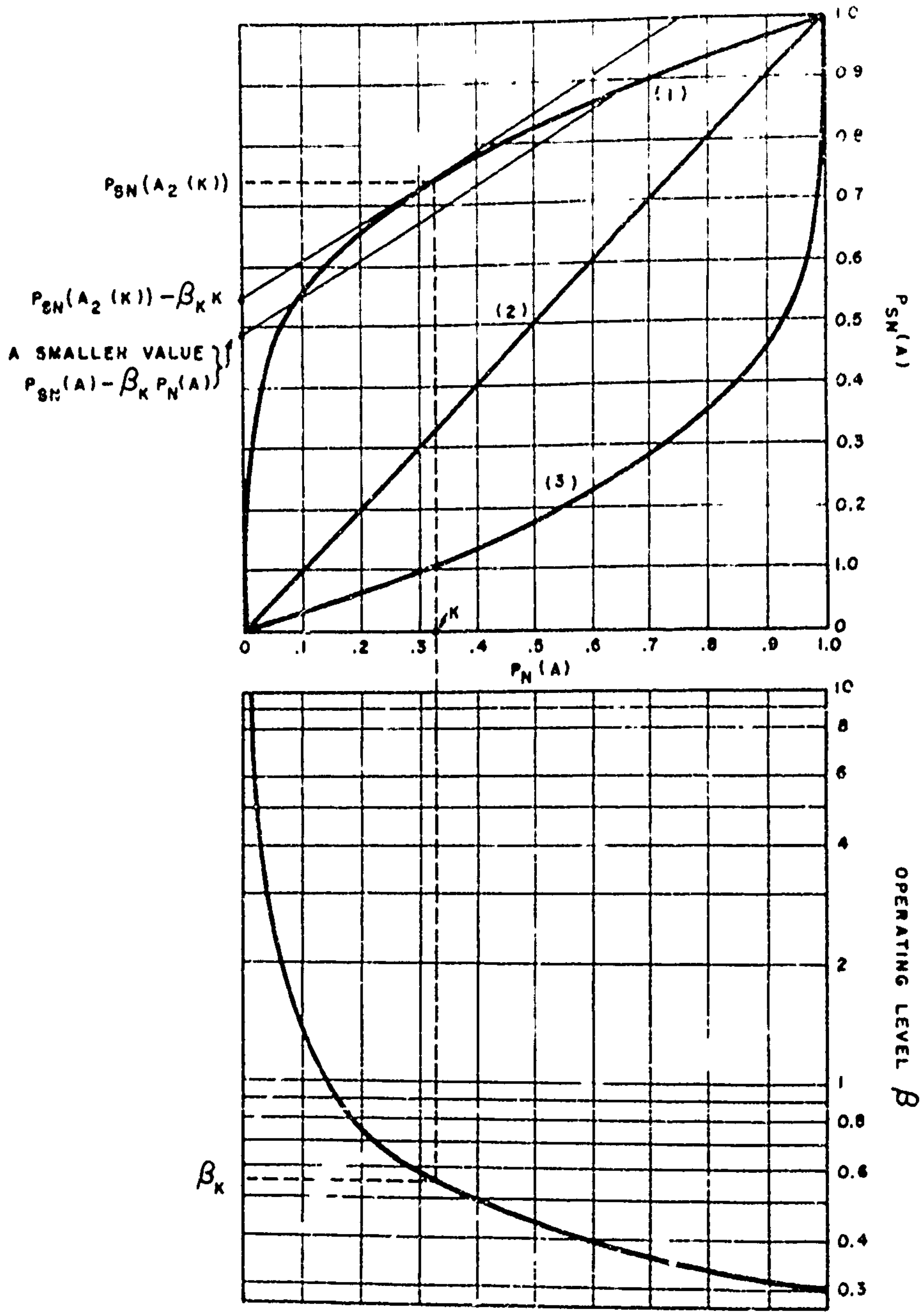


FIG. 13

TYPICAL RECEIVER OPERATING CHARACTERISTIC is made up of all inputs which have likelihood ratio greater than or equal to β. The relationship between the first and second types of optimum criteria is graphically illustrated in Fig. 1.3. If at any point $(P_N(A), P_{SN}(A))$ on curve (1) a line is drawn with slope β, it will be tangent to the curve and will intersect the axis at the value $P_{SN}(A) - \beta P_N(A)$. This is the quantity to be maximized for the first type of optimum criterion, and if a line with the same slope is drawn through any other point on or between curves (1) and (3), it will cut the axis below the point where the tangent cuts the axis. Thus, curve (1) is not only the curve for the optimum of the type when $P_N(A)$ is bounded and $P_{SN}(A)$ maximized, but also the curve for the optimum criterion when values are placed on the operator's responses.

A non-optimum receiver can be evaluated in a given situation if its receiver operating characteristic is drawn together with that of the optimum. One receiver is better than another over a range if it is closer to the optimum than the other. In some instances the optimum curve for a given situation will nearly match another receiver's operation in the same situation except that the optimum will require less signal energy. In this case, the non-optimum receiver can be given a db rating for that situation.

Each application of the theory treated in Part II of this report is accompanied by the receiver operating characteristic of the optimum receiver.

2. MATHEMATICAL THEORY

2.1 Introduction

The method for handling the signal detectability problem mathematically is described in this section. The first step is the presentation of the appropriate mathematical description of the signals and noise. In these terms the signal detectability problem is restated in several forms discussed in Section 1 of this report. It is then shown that in each case, if the likelihood ratio can be determined for each receiver input, the problem is essentially solved. Thus the conclusion is that the receiver design problem should be treated in terms of likelihood ratio; this is the approach used in Part II.

2.2 Mathematical Description of Signals and Noise

Any receiver input, noise or signal plus noise, is a voltage which is a function of time. Thus we shall be considering a set of functions. In this report it will be assumed that the receiver input is limited to bandwidth W, and that the observation is of finite duration T. By the sampling theorem,[1] any such function is completely determined when its values at "sampling" points spaced 1/2W seconds apart through the observation interval are known. There are 2WT sampling points in all. Thus a receiver input can be considered as a point in a 2WT dimensional space, the values at the sample points being taken as coordinates. Let us call the space R.

If there is noise at the receiver input, the receiver input voltage may usually be any of an infinite number of functions, i.e., any of an infinite number of points in the 2WT dimensional space R. With Gaussian noise any point in

[1] Shannon, C. E., "Communication in the Presence of Noise," Proc. IRE, Vol. 37, p. 10, January 1949; also Appendix D of Part II.

theoretically possible. It is a matter of chance which one occurs. Thus it appears that the appropriate way to describe the noise is to give the probability density for points in the space of receiver inputs.[1] The same is true when there is signal plus noise, so that we shall deal with the space R and two probability density functions, $f_N(x)$ for the case of noise alone, and $f_{SN}(x)$ for the case of signal plus noise. Here x denotes a point of the space R.

In a practical application, information will be given about the signals as they would appear without noise at the receiver input rather than about the signal plus noise probability density. Then $f_{SN}(x)$ must be calculated from this information and the probability density function $f_N(x)$ for the noise. The noise and the signals will be assumed independent. If the signals can be described by a probability density function $f_S(x)$,

$$f_{SN}(x) = \int_R f_N(x-s)\, f_S(s)\, ds \quad , \tag{2.1}$$

where the integration is over the whole space R. The receiver input x(t) could be caused by any signal s(t), and noise x(t) - s(t). The probability density for x is the probability that both s(t) and x(t) - s(t) will occur at the same time, summed over all possible s(t).

If the signals cannot be described by a probability density function, a more general form must be used, in which the signals are described by a probability measure, P_S; the formula for this case is

$$f_{SN}(x) = \int_R f_N(x-s)\, d\, P_S(s) \quad . \tag{2.2}$$

This is what is called a Lebesgue integral, and it means essentially to average

[1]We shall assume that the probability density function exists. See Appendix A

$f_N(x-s)$ over all values of s in the whole space weighting according to the probability P_S of the points s appearing as signals.[1]

2.3 A Posteriori Probability

The approach of Woodward and Davies[2] to the signal detectability problem is to ask the operator, "What is the probability that a signal is present?" He is to give the probability, using knowledge of the receiver input, i.e., he gives the a posteriori probability.

If the probability density functions are continuous, the a posteriori probability $P_x(SN)$ can be found for any particular receiver input x. Bayes' theorem[3] is used, but not directly, since $P_{SN}(x)$ and $P_N(x)$ are both zero. Consider a small sphere U with radius r and center x. Then $P_U(SN)$ can be obtained by Bayes' theorem, and $P_x(SN)$ can be defined as the

$$P_x(SN) = \lim_{r \to 0} P_U(SN) \quad . \tag{2.3}$$

Denote by $P(SN \cdot U)$ the probability that signal plus noise will be present and the receiver output will be in U. Then

$$P(SN \cdot U) = P(SN) \cdot P_{SN}(U) = P_U(SN) \cdot P(U) \tag{2.4}$$

and

$$P(U) = P_{SN}(U)\, P(SN) + P_N(U)\, (1 - P(SN)) \tag{2.5}$$

Solving for $P_U(SN)$,

$$P_U(SN) = \frac{P(SN)\, P_{SN}(U)}{P(SN)\, P_{SN}(U) + [1 - P(SN)]\, P_N(U)}$$

$$= \frac{P(SN)\, \frac{P_{SN}(U)}{P_N(U)}}{P(SN)\, \frac{P_{SN}(U)}{P_N(U)} + (1 - P(SN))} \tag{2.6}$$

---

[1] Cramér, Ref. 14, pp. 62, 188. [2] Woodward and Davies, Ref. 3. [3] Cramér, Ref. 14, p. ...

By the definition of probability density function,

$$P_{SN}(U) = \int_U f_{SN}(x)\, dx$$

$$P_N(U) = \int_U f_N(x)\, dx \quad , \tag{2.7}$$

where the integral is really a multiple integral over the volume of the sphere U in the n-dimensional space. Then

$$\frac{P_{SN}(U)}{P_N(U)} = \frac{\int_U f_{SN}(x)\, dx}{\int_U f_N(x)\, dx} \quad , \tag{2.8}$$

and if $f_{SN}(x)$ and $f_N(x)$ are continuous,

$$\lim_{r \to 0} \frac{P_{SN}(U)}{P_N(U)} = \frac{f_{SN}(x)}{f_N(x)} = \ell(x) \quad . \tag{2.9}$$

The ratio of probability densities $f_{SN}(x)/f_N(x) = \ell(x)$ is called the likelihood ratio. It follows that

$$P_x(SN) = \lim_{r \to 0} P_U(SN) = \frac{P(SN)\,\ell(x)}{P(SN)\,\ell(x) + [1 - P(SN)]} \tag{2.10}$$

This is the existence probability as defined by Woodward and Davies.[1] Notice that the likelihood ratio $\ell(x)$ is the all-important quantity. $P_x(SN)$ is a simple monotone increasing function of the likelihood ratio. Therefore if $P(SN)$ is known and if the receiver produces $\ell(x)$, a calibration will convert this to $P_x(SN)$.

2.4 Criteria and the Optimum Criteria

2.4.1 Definitions. Suppose the operator is only required to guess whether or not there is a signal present. For certain receiver inputs he will guess there is a signal present. These receiver inputs form a subset of

---

[1] Ref. 3.

the space R of all possible receiver inputs. Let us call this subset the criterion and denote it by A. That is, a point x is in the criterion A if the operator will say there is a signal present when x occurs as receiver input

It will be convenient to have a symbol for each of the two types of optimum criteria described in Section 1.4. The first type will be denoted by $A_1(\beta)$; that is, $A_1(\beta)$ is any subset of R such that for fixed $\beta \geq 0$,

$$P_{SN}\left[A_1(\beta)\right] - \beta P_N \left[A_1(\beta)\right] \text{ is maximum.} \qquad (2.11)$$

The second type will be denoted by $A_2(k)$; that is, $A_2(k)$ is any subset of R such that

$$\begin{aligned} &P_N (A_2(k)) \leq k \quad ; \quad \text{and} \\ &P_{SN} (A_2(k)) \quad \text{is maximum.} \end{aligned} \qquad (2.12)$$

The likelihood ratio $\mathcal{L}(x)$, which is defined as ratio of the probability density functions, $f_{SN}(x)/f_N(x)$ plays an important role in the following discussion. It is a measure of how much more likely the receiver input is to be if there is signal plus noise than if there is noise alone.

2.4.2 Theorems on Optimum Criteria. The optimum criterion is closely related to the likelihood ratio. For the first type of criterion the connection is given by the following theorems.

Theorem 1: Denote by A the set of points for which the likelihood ratio $\mathcal{L}(x) \geq \beta$. Then A is an optimum criterion $A_1(\beta)$.

Proof: The condition that A be an optimum criterion $A_1(\beta)$ is that $P_{SN}(A) - \beta P_N(A)$ is maximum; i.e., for any other set B of receiver inputs $P_{SN}(A) - \beta P_N(A) \geq P_{SN}(B) - \beta P_N(B)$.

$$P_{SN}(A) - \beta P_N(A) = \int_A f_{SN}(x)\,dx - \beta \int_A f_N(x)\,dx$$
$$= \int_A \left[f_{SN}(x) - \beta f_N(x) \right] dx \quad , \tag{2.13}$$

where the integration is over the set A, and so is really a multiple integral over a part of the space R which has 2WT dimensions. Let B be any set different from A. Denote by A-B the set of points which are in A and not in B, by B-A the set of points which are in B but not in A, and by $B \cap A$ the set of points which belong to both A and B. Then since A is the union of A-B and $A \cap B$, and A-B and $A \cap B$ have no points in common,

$$P_{SN}(A) - \beta P_N(A) = \int_A \left[f_{SN}(x) - \beta f_N(x) \right] dx$$
$$= \int_{A \cap B} \left[f_{SN}(x) - \beta f_N(x) \right] dx \tag{2.14}$$
$$+ \int_{A-B} \left[f_{SN}(x) - \beta f_N(x) \right] dx$$

Likewise

$$P_{SN}(B) - \beta P_N(B) = \int_{A \cap B} \left[f_{SN}(x) - \beta f_N(x) \right] dx$$
$$+ \int_{B-A} \left[f_{SN}(x) - \beta f_N(x) \right] dx \tag{2.15}$$

Thus

$$P_{SN}(A) - \beta P_N(A) - \left[P_{SN}(B) - P_N(B) \right] =$$
$$\int_{A-B} \left[f_{SN}(X) - \beta f_N(x) \right] dx - \int_{B-A} \left[f_{SN}(x) - \beta f_N(x) \right] dx \tag{2.16}$$

The points in A-B are in A, and so for them $f_{SN}(x)/f_N(x) = \mathcal{L}(x) \geq \beta$, so that $f_{SN}(x) - \beta f_N(x) \geq 0$, and the first integral in Eq. (2.16) is not less than zero. The points in the set B-A are not in A, so $f_{SN}(x)/f_N(x) < \beta$, and the second integral in Eq. (2.16) is no greater than zero. Thus

$$P_{SN}(A) - \beta P_N(A) \geq P_{SN}(B) - \beta P_N(B) \quad , \tag{2.17}$$

$P_{SN}(A) - \beta P_N(A)$ is a maximum, and A is an optimum criterion $A_1(\beta)$

There is not a unique optimum criterion $A_1(\beta)$. In the first place "optimum" was defined in terms of probability. Thus a change in $A_1(\beta)$ which would not change $P_{SN}\left[A_1(\beta)\right]$ or $P_N\left[A_1(\beta)\right]$ would result in an equally good criterion. Such a change might consist of adding or taking out a single point, a finite number of points, or generally any set of probability zero.[1] More insight into the uniqueness is given by the following theorem.

Theorem 2: If A is an optimum criterion $A_1(\beta)$, then the set of points in A for which $\mathcal{L}(x) < \beta$ has probability zero, and the set of points not in A for which $\mathcal{L}(x) > \beta$ has probability zero.

Proof: We will show that any criterion which does not have these two properties is not an optimum criterion. Consider any criterion B with a subset C, of non-zero probability, such that the likelihood ratio of each point in C is less than β. There is a positive number ϵ and a subset C_ϵ of C, having non-zero probability, such that $\mathcal{L}(x) \leq \beta - \epsilon$ for the points in C_ϵ. If this were not true, then for any positive small number ϵ, the subset C_ϵ would have probability zero. These subsets C_ϵ are monotone, that is,

[1] A set E will be said to have probability zero if both $P_{SN}(E)$ and $P_N(E)$ are zero.

if $\epsilon_2 < \epsilon_1$, then C_{ϵ_2} contains C_{ϵ_1}, and, since C contains no points with likelihood ratio equal to β, the union of all C_ϵ is C itself, and would have probability zero.[1]

As in Eq. (2.14),

$$P_{SN}(C_\epsilon) - \beta P_N(C_\epsilon) = \int_{C_\epsilon} \left[f_{SN}(x) - \beta f_N(x)\right] dx = \int_{C_\epsilon} f_N(x)\left[\ell(x) - \beta\right] dx$$

and since $\ell(x) \leq \beta - \epsilon$ or $\ell(x) - \beta \leq -\epsilon$,

$$P_{SN}(C_\epsilon) - \beta P_N(C_\epsilon) \leq -\epsilon \int_{C_\epsilon} f_N(x)\, dx = -\epsilon P_N(C_\epsilon) \quad . \tag{2.19}$$

Therefore, if $P_N(C_\epsilon) > 0$,

$$P_{SN}(C_\epsilon) - \beta P_N(C_\epsilon) < 0 \quad . \tag{2.20}$$

But C_ϵ is a subset of A, and therefore

$$P_{SN}(B - C_\epsilon) - \beta P_N(B - C_\epsilon) > P_{SN}(B) - \beta P_N(B) \quad , \tag{2.21}$$

and B is not an $A_1(\beta)$. It can be shown in an analogous manner that if there is a set D of non-zero measure outside of criterion B such that $\ell(x) > \beta$ in D, then there is a subset D_ϵ of D such that

$$P_{SN}(D_\epsilon) - \beta P_N(D_\epsilon) > 0 \tag{2.22}$$

and therefore

$$P_{SN}(B \cup D_\epsilon) - \beta P_N(B \cup D_\epsilon) > P_{SN}(B) - \beta P_N(B) \quad , \tag{2.23}$$

and B is not an $A_1(B)$.

---

[1] Cramér, Ref. 14, p. 50, Eq. 6.2.3; and p. 77, paragraph 8.2.

This theorem says nothing about the points for which $\mathcal{L}(x) = \beta$. It is not hard to show that $P_{SN}(A) - \beta\, P_N(A)$ is not affected by including or excluding points where $\mathcal{L}(x) = \beta$. Thus a criterion $A_1(\beta)$ must include all points for which $\mathcal{L}(x) > \beta$ (except perhaps a set of probability zero), none of the points where $\mathcal{L}(x) < \beta$ (except perhaps a set of probability zero), and it may or may not include a point for which $\mathcal{L}(x) = \beta$.

In the most general case, when the noise is Gaussian, the following two theorems show the uniqueness of $A_1(\beta)$.

Theorem 3: If the probability density function for noise alone, $f_N(x)$, is an analytic function, then the set of points for which $\mathcal{L}(x) = \beta$ has probability zero.[1]

A function is said to be analytic if it is analytic in the ordinary sense when considered as a function of each single coordinate. The proof of the theorem is quite involved, and so it is given in Appendix B.

Theorem 4 follows immediately from Theorem 2 and Theorem 3.

Theorem 4. If the probability density function for noise alone $f_N(x)$ is analytic, any two optimum criteria $A_1(\beta)$ can differ only by a set of probability zero.

Now let us turn to the second type of optimum criterion.

Theorem 5: Let A be a set such that if x is in A, the likelihood ratio $\mathcal{L}(x) \geq \beta$, while if x is not in A, $\mathcal{L}(x) \leq \beta$. Then if $P_N(A) = k$, A is an optimum criterion $A_2(k)$.

> Proof: An optimum criterion $A_2(k)$ must satisfy the conditions $P_N(A) \leq k$, and $P_{SN}(A)$ is maximum. The first is satisfied by hypothesis. Suppose B is any other set such that $P_N(B) \leq k$. Denote by A-B the set of points in A which are not in B, by B-A

[1] A little more is needed in the hypothesis for Theorem 3 than that $f_N(x)$ is analytic. See Appendix B.

the set of points in B which are not in A, and by $B \cap A$ the set of points common to B and A. Since A is the union of A-B and $A \cap B$, and since A-B and $A \cap B$ have no points in common,

$$P_N(A) = \int_A f_N(x)\,dx = \int_{A-B} f_N(x)\,dx + \int_{A \cap B} f_N(x)\,dx$$
$$= P_N(A-B) + P_N(A \cap B) = k \quad . \tag{2.24}$$

Likewise

$$P_N(B) = P_N(B-A) + P_N(A \cap B) \leq k, \tag{2.25}$$

and thus

$$P_N(A-B) \geq P_N(B-A) \quad . \tag{2.26}$$

Also,

$$P_{SN}(B-A) = \int_{B-A} f_{SN}(x)\,dx \quad , \tag{2.27}$$

and since any point x in B-A is not in A, $\ell(x) = \frac{f_{SN}(x)}{f_N(x)} \leq \beta$ and

$$P_{SN}(B-A) = \int_{B-A} \frac{f_{SN}(x)}{f_N(x)}\, f_N(x)\,dx \quad \beta \int_{B-A} f_N(x)\,dx \quad ,$$

or

$$P_{SN}(B-A) \leq \beta\, P_N(B-A) \quad . \tag{2.28}$$

Likewise

$$P_{SN}(A-B) \geq \beta\, P_N(A-B) \quad . \tag{2.29}$$

Collecting Eqs. (2.26), (2.28), and (2.29),

$$P_{SN}(B-A) \leq \beta\, P_N(B-A) \leq \beta\, P_N(A-B) \leq P_{SN}(A-B) \quad . \tag{2.30}$$

As in Eq. (2.24),

$$P_{SN}(A) = \int_A f_{SN}(x)\,dx = \int_{A-B} f_{SN}(x)\,dx + \int_{A\cap B} f_{SN}(x)\,dx$$
$$= P_{SN}(A-B) + P_{SN}(A\cap B) \quad , \tag{2.31}$$

and

$$P_{SN}(B) = P_{SN}(B-A) + P_{SN}(A\cap B) \quad . \tag{2.32}$$

Therefore

$$P_{SN}(A) - P_{SN}(B) = P_{SN}(A-B) - P_{SN}(B-A). \tag{2.33}$$

From Eqs. (2.30) and (2.33) it follows that

$$P_{SN}(A) \geq P_{SN}(B) \quad , \tag{2.34}$$

and $P_{SN}(A)$ is a maximum.

It follows from Theorem 5 that every optimum of the first type, $A_1(\beta)$, is an optimum of the second type. More precisely, if set A is an optimum of the first type it is associated with the fixed β for which it is an $A_1(\beta)$. By Theorem 2, the likelihood ratio in A is not less than β, and outside A the likelihood ratio is not greater than β, except on a set of probability zero. But the introduction or omission of such a set has no effect on $P_{SN}(A)$ or $P_N(A)$. Since $P_N(A)$ has some value, call it a; A will be an $A_2(a)$ by Theorem 5.

Theorem 6: For every k between 0 and 1 there is an optimum criterion of the first type A_k, such that $P_N(A_k) = k$.

Proof: For each value β we consider the maximal $A_1(\beta)$; by Theorem 2 this is the set consisting of all points of likelihood ratio not less than β:

$$M_\beta = \left\{ x \mid \ell(x) \geq \beta \right\} \quad . \tag{2.3?}$$

Now if for k there is a β such that $P_N(M_\beta) = k$, then because M_β is an $A_1(\beta)$ the proof is complete.

Next we point out that M_0 is the whole space R and M_∞ is the empty set, and therefore $P_N(M_0) = 1$ and $P_N(M_\infty) = 0$. For any value of k, if there is no M_β such that $P_N(M_\beta) = k$, let $\beta^* = \min \left\{ \beta \,\middle|\, P_N(M_\beta) \geq k \right\} = \text{g}\ell\text{b} \left\{ \beta \,\middle|\, P_N(M_\beta) < k \right\}$ that is, $P_N(M_\beta^*) > k$ and if $\beta > \beta^*$, $P_N(M_\beta) < k$. Thus the jump in P_N is due to those points in M_β^* for which $\ell(x) = \beta^*$.

Because the probability density functions exist, every point has probability zero and therefore there is a subset S of these points with $\ell(x) = \beta^*$ for which $P_N = P_N(M_\beta^*) - k$. This is shown in Appendix B (Lemma 4).

Removing this subset from M_β^*, $\quad P_N(M_\beta^* - S) = k \quad .$ (2.36)

Because $M_\beta^* - S$ satisfies Theorem 1, it is an $A_1(\beta^*)$. Of course, by Theorem 5, it is an $A_2(k)$ also.

The following theorem completes this circle of proof.

<u>Theorem 7</u>: For any k there is a β_k such that every $A_2(k)$ is an $A_1(\beta_k)$.

<u>Proof</u>: Let A be any $A_2(k)$.

By Theorem 6 there exists a β_k and an $A_1(\beta_k)$, which we will denote by A^*, such that $P_N(A^*) = k$. Then by Theorem 5, A^* is also an $A_2(k)$, and hence for both A and A^*, P_{SN} is maximum and $P_N \leqq k$. Therefore

$$P_{SN}(A^*) = P_{SN}(A) \tag{2.37}$$

$$P_N(A^*) = k \geqq P_N(A) \quad , \tag{2.38}$$

Multiplying Eq. (2.38) by $-\beta_k$ and adding gives

$$P_{SN}(A^*) - \beta_k P_N(A^*) \leqq P_{SN}(A) - \beta_k P_N(A) \quad . \qquad (\quad)$$

Since A^* maximizes this expression, the equality must hold, and A is also an $A_1(\beta_k)$.

In summary, these theorems show that β can be written as a multivalued function of k and that k can be written as a multivalued function of β. These relations can be sharpened somewhat.

Theorem 8: Let $a < b$ be two values taken on by $\mathcal{L}(x)$. If no set of the form $\left\{x \mid \mathcal{L}_1 < \mathcal{L}(x) < \mathcal{L}_2\right\}$ for $a \le \mathcal{L}_1 < \mathcal{L}_2 \le b$ has probability zero, then β_k is a single valued function of k on some interval I, with $a \le \beta_k \le b$, and $d\, P_{SN}(A_1(\beta_k))/dk$ exists and equals β_k for every k in I.

Proof: 1) In general, if a function is monotone on an interval and its range of values is also an interval, then it is continuous. If it were not, then at some point the left and right hand limits would be unequal, which would introduce a gap in the range of values, contradicting the hypotheses.

2) If $\beta_{k_1} > \beta_{k_2}$ and if the interval from β_{k_1} to β_{k_2} contains a subinterval of $\left[a, b\right]$ of length greater than zero, then $k_2 > k_1$. There are, by Theorem 6, criteria of the first type A_i (for i = 1, 2), which, by Theorem 2, may be chosen so that A_i contains all points for which $\mathcal{L}(x) > \beta_{k_i}$ and no points for which $\mathcal{L}(x) < \beta_{k_i}$. Also $P_N(A_i) = k_i$, by Theorem 5. By applying P_N to the equation $A_2 = A_1 \cup (A_2 - A_1)$, one obtains $k_2 = k_1 + P_N(A_2 - A_1)$. If $P_N(A_2 - A_1) = 0$, then from Eqs. 2.7 and the fact that $\mathcal{L}(x)$ is bounded on $A_2 - A_1$, it follows that $P_{SN}(A_2 - A_1) = 0$ also. But, by hypotheses, $A_1 - A_2$ cannot have probability zero. Hence $k_2 > k_1$.

3) Let I be the set of points k for which at least one β_k is in the open interval from a to b, and let β_k denote the possibly multivalued function defined on I. Then 2) says that β_k is both single valued and monotone, and Theorems 1 and 6 imply that the range of values of β_k is the interval from a to b. Hence I is an interval, for if it were not, there would exist three values $k_1 < k_2 < k_3$ with only the middle one not in I. Then $\beta_{k_1} < \beta_{k_2} < \beta_{k_3}$ and β_{k_2} would not be in the interval from a to b, yet the other two would be--a contradiction. Thus 1) can be applied to β_k and β_k is therefore continuous on I.

4) To form the derivative, let

$$\begin{aligned} D &= A_1(\beta_k) - A_1(\beta_{k_o}) \text{ if } \beta_k \le \beta_{k_o} \\ &= A_1(\beta_{k_o}) - A_1(\beta_k) \text{ if } \beta_k \ge \beta_{k_o} \quad . \end{aligned} \tag{2.42}$$

Then

$$\lim_{k \to k_o^+} \frac{P_{SN}(A_1(\beta_k)) - P_{SN}(A_1(\beta_{k_o}))}{k - k_o} = \lim_{k \to k_o +} \frac{P_{SN}(D)}{k - k_o} \tag{2.43}$$

Since $k \ge k_o$, $\beta_k \le \beta_{k_o}$, and in D, $\beta_k \le \ell(x) \le \beta_{k_o}$, $\beta_k f_N(x) \le f_{SN}(x) \le \beta_{k_o} f_N(x)$. But

$$P_{SN}(D) = \int_D f_{SN}(x)\, dx = \int_D \ell(x)\, f_N(x)\, dx \tag{2.44}$$

and

$$P_N(D) = k - k_o = \int_D f_N(x)\, dx \tag{2.45}$$

and therefore $\beta_k P_N(D) \le P_{SN}(D) \le \beta_{k_o} P_N(D)$. Similarly if $k \le k_o$ $\beta_{k_o} P_N(D) \le P_{SN}(D) \le \beta_k P_N(D)$. Thus

$$\lim_{k \to k_o} \frac{P_{SN}(D)}{k - k_o} = \beta_{k_o} , \qquad (2.46)$$

by virtue of the result that β_k is a continuous function of k.

2.5 Evaluation of Optimum Receivers

2.5.1 Introduction. This section treats the problem of determining how well a given receiver will perform its task of detecting signals. For the criterion type receiver, the probability of false alarm if no signal is sent, $P_N(A)$, and the probability of detection if a signal is sent, $P_{SN}(A)$, give a good measure of receiver performance. For the a posteriori probability type receivers, the average or mean a posteriori probability with signal plus noise and with noise alone describe the receiver's ability to discriminate between signal plus noise and noise alone.

2.5.2 Evaluation of Criterion Type Receivers. For simplicity, let us restrict this discussion to the case in which the probability density function for noise alone, $f_N(x)$ is analytic.

Denote by $F_{SN}(\beta)$ the probability that the likelihood ratio $\ell(x)$ is equal to or greater than β if there is signal plus noise, and similarly, let $F_N(\beta)$ be the probability that $\ell(x)$ is equal to or greater than β if there is noise alone. These are the complimentary distribution functions for $\ell(x)$. Then for any $A_1(\beta)$,

$$P_{SN}\ (A_1(\beta)) = F_{SN}(\beta), \text{ and} \qquad (2.47)$$

$$P_N\ (A_1(\beta)) = F_N(\beta), \qquad (2$$

because the set of points for which $\ell(x) \geq \beta$, and differs from any $A_1(\beta)$ only by a set of probability zero (Theorem 4). By Theorem 7, every $A_2(k)$ is an $A_1(\beta)$. The β_k corresponding to k can be found from Eq. (2.48)

$$F_N(A_1(\beta_k)) = F_N(\beta_k) = k \quad . \tag{2.49}$$

Then

$$F_{SN}(A_2(k)) = F_{SN}(\beta_k) \quad . \tag{2.50}$$

Thus, if the distribution functions $F_{SN}(\beta)$ and $F_N(\beta)$ are known, any criterion type receiver can be evaluated.

It turns out that not both $F_{SN}(\beta)$ and $F_N(\beta)$ are necessary. Theorem 8 states that

$$\frac{d\,F_{SN}(\beta)}{d\,F_N(\beta)} = \beta \quad , \tag{2.51}$$

since $F_{SN}(A_1(\beta_k)) = F_{SN}(\beta_k)$, and $k = F_N(\beta_k)$. Thus, if $F_N(\beta)$ is known, $F_{SN}(\beta)$ can be found by integrating Eq. (2.51).[1]

$$F_{SN}(\beta) = -\int_{\beta}^{\infty} y \, d\,F_N(y) \quad . \tag{2.52}$$

As an alternative, $F_{SN}(\beta)$ might be given as a function of $F_N(\beta)$; this is the receiver operating characteristic graph. Then β can be found from Eq. (2.51); i.e., β is the slope of the graph.

[1] The change in sign is because the functions $F_{SN}(\beta)$ and $F_N(\beta)$ are complimentary distribution functions. If the density function associated with $F_N(\beta)$ is $g(\beta)$, then $\frac{d\,F_N(\beta)}{d\beta} = -\,g(\beta)$ and $F_{SN}(\beta) = \int_{\beta}^{\infty} g(\beta)\,d\,\beta$.

A corollary of Theorem 8 is the following: The nth moment of the distribution for noise alone is the (n-1)st moment of the signal plus noise distribution.

$$\int_{-\infty}^{\infty} y^n \, d\, F_N(y) = \int_{-\infty}^{\infty} y^{n-1} \, (y \, d\, F_N(y)) = \int_{-\infty}^{\infty} y^{n-1} \, d\, F_{SN}(y) \qquad (2.53)$$

As an example of the application of this corollary, note that the mean value of likelihood ratio with noise alone is always unity. If the variance with noise alone is σ_N^2, the second moment of $F_N(\beta)$ is $1 + \sigma_N^2$; then the mean of the signal plus noise distribution is $1 + \sigma_N^2$, and the difference of the means is σ_N^2. For detection corresponding roughly to Fig. 2.1, the difference of the means of the two distributions must be of the order of the standard deviation of the distributions, so that

$$\sigma_N^2 \approx \sigma_N \, , \qquad (2.54)$$

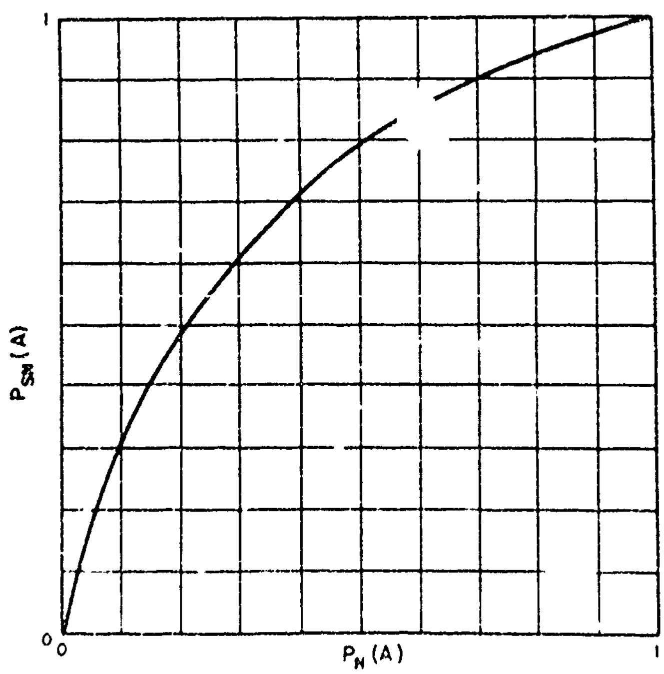


FIG. 2.1
RECEIVER OPERATING CHARACTERISTIC
For $\sigma_N^2 = 1$.

or the variance of the distribution with noise alone must be of the order of unity. For better detection, σ_N^2 must be greater.

2.5.3 Evaluation of A Posteriori Probability Woodward and Davies Type Receivers. Davies proposes the mean a posteriori probability as a measure of the efficiency of a receiver. The mean a posteriori probability is defined as:

$$\mu_{SN}(P_x(SN)) = \int_R P_x(SN)\, f_{SN}(x)\, dx \tag{2.55}$$

$$\mu_N(P_x(SN)) = \int_R P_x(SN)\, f_N(x)\, dx \tag{2.56}$$

These can be evaluated if the distribution functions $F_{SN}(\beta)$ and $F_N(\beta)$ for likelihood ratio are known. Since

$$P_x(SN) = \frac{P(SN)\ \ell(x)}{P(SN)\ \ell(x) + 1 - P(SN)} , \tag{2.57}$$

the mean a posteriori probabilities are

$$\mu_{SN}(P_x(SN)) = \int \frac{y\, P(SN)}{y\, P(SN) + 1 - P(SN)}\, d\, F_{SN}(y), \text{ and} \tag{2.58}$$

$$\mu_N(P_x(SN)) = \int \frac{y\, P(SN)}{y\, P(SN) + 1 - P(SN)}\, d\, F_N(y) . \tag{2.59}$$

Davies presents the formula

$$\mu_{SN}\left[P_x(SN)\right] + \frac{1 - P(SN)}{P(SN)}\ \mu_N\left[P_x(SN)\right] = 1 , \tag{2.60}$$

which enables one to calculate easily either one of the mean a posteriori probabilities once the other has been calculated.

2.6 Conclusions

It is possible to combine the most common statistical approaches to the theory of signal detectability into one general theory. In this theory likelihood ratio plays the central role: the result of the theory is that a receiver built so that its output is likelihood ratio can be adapted easily to accomplish the task specified in any of the well-known approaches to signal detectability. If the probability distribution of likelihood ratio is known, then the receiver reliability can be evaluated.

In Part II of this report, likelihood ratio and its distribution functions are calculated for a umber of specific cases, and the problems of receiver design are discussed.

APPENDIX A

It was assumed throughout the discussion of the criterion approach to signal detectability that for any given receiver input, the operator would always give the same response. This is certainly not the case with threshold signals and a human operator. A more realistic approach might be to assume that for any receiver input x, the operator would say with probability $\rho(x)$ that there is signal plus noise. Finding the optimum receiver would then consist of finding the optimum $\rho(x)$. This approach does not lead to any interesting new results; if $\rho(x) = 1$ on an optimum criterion and zero on its compliment, then $\rho(x)$ is optimum.

The theorems on signal detectability are proved in Section II in more general form than has yet been found necessary in an application. However, they can be generalized somewhat, and this appendix discusses some of the possibilities.

It is certainly possible to consider more general spaces of signals. Any space on which a probability measure can be defined might be used. In order to prove the theorems on optimum criteria, however, some sort of likelihood ratio seems necessary. One possibility is to assume the measure $P_N(A)$ and the random variable $\mathcal{L}(x)$ are given and to define $P_{SN}(A)$ through the integral

$$P_{SN}(A) = \int_A \mathcal{L}(x)\, d\, P_N(A) \quad . \tag{A.1}$$

The mean value of $\mathcal{L}(x)$ must be unity, of course.

If the space is a Euclidean space of finite dimension, then it is possible to define an arbitrary measure through distribution functions. These functions, being monotone, have a derivative almost everywhere, and thus affo. a means of defining likelihood ratio. For any point which has measure zero, the likelihood is the ratio of the derivatives of the distribution function for signal plus noise and for noise alone. Points which do not have measure zero can always be treated separately. There can be only a countable number of these and likelihood ratio for such a point x can be defined as

$$\ell(x) = \frac{P_{SN}(x)}{P_N(x)} \tag{A.2}$$

Any point with infinite likelihood ratio belongs in the criterion, of course, and such a point has a posteriori probability unity. Then likelihood ratio is defined except for a set of points of measure zero.

In any case where likelihood ratio is defined and satisfies Eq. (A.1), Theorems 1 and 2 can be proved. The lemma (Appendix B, Lemma 1) which is needed for the proof of Theorem 5 can be proved for any space and measure for which sets of arbitrarily small measure can be found containing each point. If this holds and likelihood ratio is defined, then Theorems 5, 6, 7, and 8 can be proved.

APPENDIX B

This appendix contains the proof of Theorem 3 and the lemma required to complete the proof of Theorem 6. It is convenient to prove three lemmas from which Theorem 3 will follow directly.

Lemma 1: Let S be a sphere (i.e., the set of all points whose distance to a fixed point is less than or equal to a fixed positive number) in n-dimensional Euclidean space E^n. Let $f(x)$ be a continuous real function defined on S. Then the graph $G = \{[x, f(x)]\}$ of $f(x)$ in E^{n+1} has (n+1)-measure zero.

Proof: Let the volume (the n-measure) of S be V. Since $f(x)$ is uniformly continuous on S, for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever the distance between x_1 and x_2 is less than δ it follows that $|f(x_1) - f(x_2)| < \epsilon/4V$.

Moreover, for each $\delta > 0$ there is a decomposition of E^n into pairwise disjoint congruent n-dimensional cubes each with its greatest diagonal of length less than $\delta/2$. This decomposition may be chosen so that, if $\{C_i\}$ $i = 1, 2, \ldots, k$ are the cubes that touch S, then

$$\sum_i \text{(volume } C_i) < 2V \quad . \tag{B.1}$$

Thus $I_i = f(C_i)$ is an interval of length less than $2(\epsilon/4V) = \epsilon/2V$.

Now, let C_i^* be the (n+1)-cube formed by the Cartesian product $C_i \times I_i$; by construction, the graph G is covered by the (n+1)-cubes C_i^*. Also

$$\sum_i \left[\text{(n+1)-volume } C_i^*\right] \leq \sum_i \left[\text{(n)-volume } C_i\right] \epsilon/2V \leq 2V \cdot \epsilon/2V = \epsilon \quad . \tag{B.2}$$

Thus for each $\epsilon > 0$ there is a covering of G by (n+1)-cubes whose total (n+1)-volume is less than ϵ. This means (n+1)-measure of G is zero.

Lemma 2: Let D be an open set in Euclidean n-dimensional space E^n and $f(x)$ a real function defined for all points x in D which has continuous partial derivatives of all orders such that at each point x in D at least one partial derivative (of any order) does not vanish. Then, if b is some value taken on by f, the set $f^{-1}(b)$ of all points x such that $f(x) = b$ has n-measure zero.

Proof: A point x in D is said to have "order zero" if some first order derivative of f does not vanish at x; x has "order r" (r a positive integer) if all partial derivatives of f of order $\leq r$ vanish at x, but at least one partial derivative of f of order r+1 does not vanish at x. By the hypotheses, every point of D has finite order.

For each integer $r \geq 0$ let C_r be the set of points in $f^{-1}(b)$ of order r; then $f^{-1}(b) = \bigcup_{r=0}^{\infty} C_r$. The theorem is proved if it is shown that the n-measure of C_r is zero for each r. This will be done in two steps.

I. At each point x^* in C_r, there is a sphere $S(x^*)$ centered at x^* such that $S(x^*) \cap C_r$ has n-measure zero.

II. There is a countable collection $\{S(x^i)\}$, $i = 1, 2, \ldots$, of such spheres such that C_r is contained in the union $\bigcup_{i=1}^{\infty} S(x^i)$.

Steps I and II together show that n-measure of C_r is zero because

$$0 \leq \text{n-measure } C_r \leq \sum_{i=1}^{\infty} \text{n-measure } \left[S(x^i) \cap C_r\right] = 0 \quad . \tag{B.3}$$

Step II is an application of the Lindelöff theorem which asserts that every collection of spheres contains a countable subcollection whose union is equal to the union of all the original spheres.

The proof of I follows:

Since x° is of order r, one of the derivatives of order r of $f(x)$, say $\omega(x)$, has a first order derivative which does not vanish at x° By a change in notation, this can be written as: $\frac{\partial \omega}{\partial x_n} = \omega_n$ does not vanish at x°

$x^\circ = (x^\circ_1, \ldots, x^\circ_n)$. The implicit function theorem can then be applied to ω, yielding these results:

1) there is a sphere $S(x^\circ)$ centered at x° and contained in D.

2) writing π for the projection of $S(x^\circ)$ onto the $x_1, \ldots, x_{n-1}$ "coordinate plane," π is an (n-1) sphere. There is a real valued continuous function $X(x_1, \ldots, x_{n-1})$ defined on π whose graph $G = \left\{\left[x_1, \ldots, x_{n-1}, X(x_1, \ldots, x_{n-1})\right]\right\}$ is the set of all points x in $S(x^\circ)$ such that $\omega(x^\circ) = \omega(x)$; that is $G = S(x^\circ) \cap \omega^{-1}\left[\omega(x^\circ)\right]$.

Note: 2) says that, in particular, $\omega\left[x_1, \ldots, x_{n-1}, X(x_1, \ldots, x_{n-1})\right] = \omega(x^\circ)$. This is the usual way of stating the theorem.

By Lemma 1, the n-measure of G is zero. Thus step I is proved if $S(x^\circ) \cap C_r \subset G$.

Case 1: $r = 0$. If x is in $S(x^\circ) \cap C_o$, then x is of order $r = 0$ and $f(x) = f(x^\circ)$. But in this case ω must have been chosen to be f, so $\omega(x) = \omega(x^\circ)$, which implies that x is in G.

Case 2: $r > 0$. If x is in $S(x^\circ) \cap C_r$, then x is of order r, which means that in particular all r-order partials of f vanish at x. Hence $\omega(x) = 0$. Also, by the same argument $\omega(x^\circ) = 0$, and $\omega(x) = \omega(x^\circ)$ implies that x is in G. This completes the proof of Lemma 2.

<u>Lemma 3</u>: If $f_N(x_1, x_2, \ldots, x_n)$ is an analytic function defined on n-dimensional Euclidean space E^n, and if $P(S_1, S_2, \ldots, S_n)$ is a probability measure on E^n such that there exists a bounded set in E^n whose probability is unity, then

$$f_{SN}(x_1, \ldots, x_n) = \int_{E^n} f_N(x_1-s_1, \ldots, x_n-s_n)\, dP(s_1, \ldots, s_n) \qquad \text{(B.4)}$$

exists and is analytic.

Proof: Let B be a bounded set such that $P(B) = 1$. Then $\bar{B}$, the closure of B, is such a set also; it is certainly bounded, and it can be assigned the measure unity, since

$$B \subset \bar{B} \subset E^n \text{ and } 1 = P(B) \leq P(\bar{B}) \leq P(E^n) = 1 \ . \qquad \text{(B.5)}$$

The probability of the complement of $\bar{B}$ is zero, and hence the integration can be restricted to the set $\bar{B}$ rather than to the whole of E^n.

For a fixed $(x_1, \ldots, x_n)$ and for $(s_1, \ldots, s_n)$ in $\bar{B}$, $f_N(x_1-s_1, \ldots, x_n-s_n)$ is bounded, since f_N is continuous and $\bar{B}$ is closed and bounded. The function f_N is also measurable, since it is continuous. (This assumes open sets are measurable.) Then the integral exists.[1]

The function $f_N(x_1, \ldots, x_n)$ being analytic means that $f_N(x_1, \ldots, x_n)$ is an analytic function in the ordinary sense when considered as a function of any single coordinate x_i. Let us forget about the other coordinates for the present. Then $f_N(x_i)$ has a power series expansion at each point x°_i, which converges in a neighborhood of the point $(x^\circ_i, 0)$ in the complex plane. Thus $f_N(x_i)$ can be extended for complex values of x_i in a region containing the real axis.

Formally,

$$\frac{\partial}{\partial x_i} f_{SN}(x_i) = \lim_{h \to 0} \frac{f_{SN}(x_i+h) - f_{SN}(x_i)}{h} \qquad \text{(B.6)}$$

[1]Cramér, Ref. 14, Section 5.2, p. 37.

$$= \lim_{h \to 0} \frac{1}{h} \left[\int_B f_N(x_1 - s_1, \ldots, x_i + h - s_i, \ldots, x_n - s_n)\, dP(s_1, \ldots, s_1) \right.$$
$$\left. - \int_B f_N(x_1 - s_1, \ldots, x_i s_i, \ldots, x_n - s_n)\, dP(s_1, \ldots, s_n) \right] \quad (B.7)$$

$$= \lim_{h \to 0} \int_B \frac{1}{h} \left[f_N(x_1 - s_1, \ldots, x_i + h - s_i, \ldots, x_n - s_n) \right.$$
$$\left. - f_N(x_1 - s_1, \ldots, x_i - s_i, \ldots, x_n - s_n) \quad dP(s_1, \ldots, s_n) \right] \quad (B.8)$$

$$= \int_B \lim_{h \to 0} \frac{1}{h} \left[f_N(x_1 - s_1, \ldots, x_i + h - s_n, \ldots, x_n - s_n) \right.$$
$$\left. - f_N(x_1 - s_1, \ldots, x_i, \ldots, s_i - x_n - s_n) \quad dP(\ _1, \ldots, s_n) \right] \quad (B.9)$$

$$= \int_B \frac{\partial f_N}{\partial x_i} \quad dP(s_1, \ldots, s_n) \quad . \quad (B.10)$$

The only question now is whether or not it is permissible to interchange the order of integration and taking the limit of the difference quotient at step (B.9). This is permissible if the difference quotient converges uniformly, which turns out to be the case.

The function $f_N(x_i)$ is analytic in a domain which extends to complex values of x_i near the real axis. The function $f_N(x_i + h - s_i)$ can be considered as a function of $h - s_i$, and is analytic for complex values of $h - s_i$ in a domain containing the real axis. Since the values of $s = (s_1, \ldots, s_n)$ in B are a closed bounded set, and the values of h can certainly be bounded, the set V of values $h - s_i$ is bounded. V can also be taken as closed, and it can be chosen so that no point s_i is on its boundary. Then there will be a minimum distance $h_0 > 0$ from points s_i to the boundary of V. Consider the function

$$\psi(s_1, \ldots, s_n, h) = \frac{1}{h}\Big[f_N(x_1-s_1, x_i+h-s_i, \ldots, x_n-s_n) - f_N(x_1-s_1, \ldots, x_i-s_i, \ldots, x_n-s_n)\Big]$$

$$\text{if } h \neq 0, \text{ and}$$

$$= \frac{\partial f_N}{\partial s_i}, \qquad \text{if } h = 0 \quad ,$$

defined for $|h| \leq h_0$, and s in $\bar{B}$. ψ is continuous at every point, and it is defined for all points (h, s) with $h = u+iv$ and $s = (s_1, \ldots, s_n)$ of a compact subset of E^{n+2}. ψ is therefore uniformly continuous, and its convergence to $\frac{\partial f_N}{\partial x_i}$ as h approaches zero along <u>any</u> complex valued path is uniform <u>in</u> s. Thus the difference quotient converges uniformly.

<u>Lemma 3'</u>: Let $f_N(x_1, \ldots, x_n)$ be a function of n complex variables, and suppose that for each i, there is a domain D_i in the complex plane and a number h_0 such that the domain D_i contains all points within a distance of h_0 of the real axis, and $f_N(x_1, \ldots, x_i, \ldots, x_n)$ is an analytic function of x_i in D_i for all real values of the other coordinates. Then, if $P(s_1, \ldots, s_n)$ is a probability measure on the n-dimensional Euclidean space E^n,

$$f_{SN}(x_1, \ldots, x_n) = \int_{E^n} f_N(x_1-s_1, \ldots, x_n-s_n)\, dP(s_1, \ldots, s_n) \qquad \text{(B.11)}$$

is analytic if it exists.[1]

[1] If f_N is bounded, the integral must exist, as in the previous case.

The proof will be omitted. The idea of the proof is as follows: one must form the difference quotient for $f_{SN}(x_1, \ldots, x_n)$ for each coordinate x_i

$$\frac{1}{h} \left[f_{SN}(x_1, \ldots, x_i+h, \ldots, x_n) - f_{SN}(x_1, \ldots, x_i, \ldots, x_n) \right]$$

and show that the limit as $h \to 0$ exists, and is equal to what is obtained by differentiating under the integral sign. The space can be divided into two parts such that one will have arbitrarily small measure and contribute an arbitrarily small amount to the integrals, while the other will be closed and bounded and hence on it the order of integration and taking the limit as $h \to 0$ can be interchanged, as in Lemma 3. The domain D_i is required so that differentiation in the complex plane will be possible.

Now let us discuss Theorem 3. Suppose $f_N(x)$ is analytic, and suppose either Lemma 3 or Lemma 3' holds. Then $f_{SN}(x)$ is analytic, and their ratio

$$\ell(x) = \frac{f_{SN}(x)}{f_N(x)} ,$$

is analytic except where $f_N(x) = 0$. This is a set of measure zero, by Lemma 2. Since $\ell(x)$ is analytic, the points where $\ell(x) = \beta$ form a set of measure zero, by Lemma 2.[1] This proves Theorem 3.

<u>Theorem 3</u>: If the probability density function for noise alone, $f_N(x)$, is an analytic function, (and if either Lemma 3 or Lemma 3' holds,) then the set of points for which $\ell(x) = \beta$ has measure zero.

The restriction that Lemma 3 or Lemma 3' holds is not at all serious. If the signals have bounded energy, Lemma 3 holds. Lemma 3' would be expected to hold for most analytic probability density functions, and in particular it does hold if the noise is Gaussian.

---

[1]Note that Lebesgue measure zero implies probability zero, since the probability [illegible] through density functions.

The following lemma is needed to complete the proof of Theorem 6.

Lemma 4: Let $f(x)$ be a probability density function defined on the n-dimensional Euclidean space E^n. Denote by $P(A)$ the value of the integral $\int_A f(x)\,dx$ for all subsets A of E^n for which the integral exists. If A_o is any P-measurable set whose measure $P(A_o)$ is finite, and if $0 < \gamma < P(A_o)$, then there is a P-measurable set B_o such that $P(B_o) = \gamma$.

The following proof makes the theorem valid for any measure on any space M with the property "C" defined below.

Proof: Under the hypotheses above, the measure P has a special property relative to the space E^n.

Property "C": There is a countable class $[C_i]$, $i = 1, 2, \ldots$, of P-measurable sets such that if x is a point and $\epsilon > 0$ then there is a C_i containing x such that $P(C_i) < \epsilon$.

One can obtain such a class by choosing all (n-dimensional) spheres of rational radius centered at points whose coordinates are rational. This class is countable because the rational numbers are countable. Its members are P-measurable because $\int_A f(x)\,dx$ exists for any sphere A. That it has property "C" is a way of stating a fundamental property of integrals.

The desired set B_o will be constructed as the union of a special sequence $[D_i]$ of P-measurable sets. Define D_1 to be $C_1 \cap A_o$ if $P(C_1 \cap A_o) \leq \gamma$; otherwise define D_1 to be empty. If D_n has been defined, define $D_{n+1} = D_n \cup [C_{n+1} \cap A_o]$ if $P\{D_n \cup [C_{n+1} \cap A_o]\} \leq \gamma$; otherwise define $D_{n+1} = D_n$. Since $D_n \subset D_{n+1}$, $P(D_n) \leq P(D_{n+1}) \leq \gamma$. Hence the sequence $[P(D_n)]$ of real numbers converges. A general property of measures yields the result that $P\left[\bigcup_{n=1}^{\infty} D_n\right] = \lim_{n \to \infty} P(D_n)$. Write $B_o = \bigcup_{n=1}^{\infty} D_n$; then $P(B_o) = \lim_{n \to \infty} P(D_n) \leq \gamma$.

It remains to be shown that $P(B_o) = \gamma$. Suppose $P(B_o) < \gamma$; then writing $\epsilon = \gamma - P(B_o) > 0$, one has $P(B_o) = \gamma - \epsilon$. Since $P(B_o) < P(A_o)$, there is a point x in A_o but not in B_o. By property "C", there is some C_k containing x such that $P(C_k) < \epsilon$. Return to the definition of D_k. If $P\left\{D_{k-1} \cup \left[C_k \cap A_o\right]\right\} \leq \gamma$, then D_k was defined to be $D_{k-1} \cup \left[C_k \cap A_o\right]$. Here

$$P\left\{D_{k-1} \cup \left[C_k \cap A_o\right]\right\} \leq P(D_{k-1}) + P(C_k) \leq P(B_o) + P(C_k) \leq (\gamma - \epsilon) + \epsilon = \gamma .$$

Thus it was the case that $C_k \cap A_o \subset D_k \subset B_o$. But $C_k \cap A_o$ contains a point x not in B_o. This contradiction shows that $P(B_o)$ is actually equal to γ and not less than as was supposed.

APPENDIX C

The following theory was developed as the preparation of the text of this report neared completion. The subject matter is appropriate to this report, and so it is included.

The purpose of this material is to characterize uniformly best tests, or criteria. If there are a family of signal distributions (or hypotheses, in statistical terms), and if a criterion A is an $A_2(k)$ for each of them, then A is a uniformly best test.[1] Theorem C1 states that if all distributions in a family of signal distributions are k-equivalent, all optimum criteria are uniform best tests, and Theorem C2 states the converse.

In the first three cases considered in Part II of The Theory of Signal Detectability, the signal known exactly, the signal known except for carrier phase, and the signal a sample of white Gaussian noise, two signal distributions differing only in signal energy are k-equivalent. Thus, by Theorem C4, a signal distribution with fixed signal energy and one with the signal energy having an arbitrary distribution are k-equivalent in these three cases. These three cases have for the boundaries of their optimum criteria, planes, cylinders, and spheres, respectively. For the other cases, with more complicated criterion boundaries, k-equivalence cannot be expected when energy is changed.

Definition: If $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ and $f_N(x)$ are defined on E^n, and if there exists a set X of probability zero such that for any two points x and y in in E^n, but not in X,

$$\ell_1(x) \gtrless \ell_1(y) \text{ if and only if } \ell_2(x) \gtrless \ell_2(y) \quad ,$$

then $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ are said to yield k-equivalent distributions.

[1] Neyman and Pearson, Ref. 13.

Theorem C1: If $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ give k-equivalent distributions, then a criterion is an $A_2(k)$ for the first if and only if it is an $A_2(k)$ for the second.

Proof: Suppose A is an $A_2(k)$ for the first distribution. Then by Theorem 7, there is a β such that A is a $A_1(\beta)$. By Theorem 2, A contains all points for which $\ell(x) > \beta$ and none for which $\ell(x) < \beta$, except for a set of probability zero. Except for a set of probability zero, if x and y are any two points such that x is in A and y is not in A, then $\ell_1(x) \geqq \ell_1(y)$. By definition of k-equivalence, there is a set X of probability zero, such that if x and y are also not in X, $\ell_2(x) \geqq \ell_2(y)$. Then there must exist a number β_2 such that for any x except a set of probability zero, $\ell_2(x) \geqq \beta_2$ if x is in A and $\ell_2(x) \leqq \beta_2$ if x is not in β_2. If follows that A is an $A_1(\beta_2)$ with respect to the second distribution. Furthermore, $P_N(A) = k$, for either distribution since the probability density with noise alone is the same for both distributions. It follows by Theorem 5 that A is an $A_2(k)$ for the second distribution.

Theorem C2: If $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ lead to two distributions such that for every k, any criterion A is an $A_2(k)$ for one if and only if it is for the other also, then $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ lead to k-equivalent distributions.

Proof: Consider the family of sets A_α where $A_\alpha = \left\{x \mid \ell_1(x) \geqq \alpha\right\}$, and α takes on all rational number values greater than zero. Each A_α is an $A_2(k)$ for some k with respect to the first distribution, by Theorem 5. Then it is for the second also, by hypothesis. Each A_α is an $A_1[\beta(\alpha)]$ for some $\beta(\alpha)$, by Theorem 7. For each A_α, the set of points C_α such that x is in A_α and $\ell(x) < \beta(\alpha)$ or x is not in A_α and $\ell(x) > \beta(\alpha)$ has probability zero, by Theorem 2. Let X_1 be the union of all the sets C_α, and since each C_α has probability zero, and the rational numbers and hence the family C_α is countable, it follows the the set X_1 has proba-

Now consider the family of sets

$$A_r = \bigcap_{\text{all } \alpha < r} A_\alpha = \left\{ x \mid \ell_1(x) \geq r \right\} \tag{C.1}$$

defined for every positive real number r. Also define

$$g(r) = \ell.\text{ u. b. } \beta(\alpha) \quad \text{all } \alpha < r \tag{C.2}$$

Then for any point x not in X_1, if x is in A_r, $\ell_2(x) \geq g(r)$. Also consider the family of sets

$$A^*_r = \bigcup_{\text{all } s > r} A_s = \left\{ x \mid \ell_1(x) > r \right\} \tag{C.3}$$

defined for every positive real number r. If x is a point not in X_1, and if x is not in A^*_r,

$$\ell_2(x) \leq \underset{\text{all } r^* > r}{\text{g. } \ell.\text{ b.}} \; g(r^*) \quad . \tag{C.4}$$

For any value of r at which g(r) is continuous,

$$g(r) = \underset{\text{all } r^* > r}{\text{g. } \ell.\text{ b.}} \; g(r^*) \quad . \tag{C.5}$$

Any point x which is not in X_1 and for which $\ell_1(x) = r$ is in A_r but not in A^*r, and therefore

$$g(r) \leq \ell_2(x) \leq g(r), \text{ i.e., } \ell_2(x) = g(r) \quad . \tag{C.6}$$

Clearly g(r) is a monotone increasing function of r. It can therefore have at most a countable number of discontinuities. Let r_o denote a discontinuity in g(r) and suppose that the set of points $B = \left\{ x \mid \ell_1(x) = r_c \right\}$ has probability greater than zero. Define

$$h(r_o) = \ell.\text{ u. b. } \left\{ \beta \mid P \left\{ x \mid x \in B \text{ and } \ell_2(x) < \beta \right\} = 0 \right\}$$

$$h^*(r_o) = \text{g. } \ell.\text{ b. } \left\{ \beta \mid P \left(\left\{ x \mid x \in B \text{ and } \ell_2(x) > \beta \right\} \right) = 0 \right\} . \tag{C.7}$$

The claim is made that $h(r_o) = h^*(r_o)$. Suppose $h(r_o) \neq h^*(r_o)$. Then there exists a number γ such that $h(r_o) < \gamma < h^*(r_o)$. Define

$$C_1 = \left\{ x \mid h(r_o) \leq \ell_2(x) \leq \gamma \right\}$$

$$C_2 = \left\{ x \mid \gamma < \ell_2(x) \leq h^*(r_o) \right\} \quad . \tag{C.8}$$

Both C_1 and C_2 have probability greater than zero, by Eq. (C.7). Now consider the set $A_r - C_2$. It is an $A_2(k)$ for the first distribution, by Theorem 5. Clearly, by Theorems 7 and 2, it cannot be an $A_2(k)$ for the second distribution. The contradiction leads us to conclude that $h(r_o) = h^*(r_o)$. Then for each discontinuity r_o there exists a set of probability zero, say $S(r_o)$, such that if $\ell_1(x) = r_o$ and x is not in $S(r_o)$, $\ell_2(x) = h(r_o)$. Let $X_2 = \bigcup_{\text{all } r_o} S(r_o)$. Then X_2 has probability zero, since there are only a countable number of points of discontinuity r_o. Now define $X = X_1 \cup X_2$, X also has probability zero. Let the function $h(r)$ be defined as follows:

$$h(r) = g(r) \quad \text{if } g(.) \text{ is continuous at } r$$

$$h(r) = h(r_o) \text{ at } r = r_o, \text{ a discontinuity of } g(r). \tag{C.9}$$

The function $h(r)$ has the following properties: (1) $h(r)$ is a monotone increasing function of r, and (2) if $\ell_1(x) = r$, and x is not in X, then $\ell_2(x) = h(r)$. The first assertion is an obvious consequence of the way in which $h(r)$ is defined, and the fact that $g(r)$ is monotone. The second assertion has been shown separately first for points where g, and hence h, is continuous, Eq. (C.6), secondly for the points of discontinuity of h, in the preceding paragraph.

Now suppose x and y are not elements of X, and $\ell_1(x) \geq \ell_1(y)$. If $\ell_1(x) = r_x$ and $\ell_1(y) = r_y$, then $r_x \geq r_y$. It follows from the fact that $h(r)$ is monotone increasing that $h(r_x) \geq h(r_y)$, and since $\ell_2(x) = h(r_x)$ and

$\ell_2(y) = h(r_y)$, $\ell_2(x) \geq \ell_2(y)$. Since X has probability zero, this completes the proof.

Theorem C3. If $f_{SN}^{(1)}(x)$ is k-equivalent to $f_{SN}^{(i)}(x)$ for each value of i between 2 and n, (or between 2 and ∞), and a_i are positive real numbers such that $\sum_1^n a_i = 1$, (or $\sum_1^\infty a_i = 1$), then $f_{SN}^{(1)}(x)$ and $\sum_1^n a_i f_{SN}^{(i)}(x)$, (or $\sum_1^\infty c_i f_{SN}^{(i)}(x)$) yield k-equivalent distributions.

The set X (in the definition of k-equivalence) for the distribution given by the sum is taken as the union of the sets X for the individual distributions. Then the proof is obvious.

Theorem C4: If $f_{SN}^{(\alpha)}(x)$ is a continuous function of α in an interval $[a, b]$, if for any two numbers α_1 and α_2, $f_{SN}^{(\alpha_1)}(x)$ is k-equivalent to $f_{SN}^{(\alpha_2)}(x)$, and if $F(\alpha)$ is a monotone function which is zero at the left end of the interval and 1 at the right end of the interval, then

$$\int_a^b f_{SN}^{(\alpha)}(x)\, dF(\alpha)$$

is k-equivalent to any $f_{SN}^{(\alpha)}(x)$.

Proof: Choose any α_o in the interval $[a, b]$. Then for each rational value of α in the interval $[a, b]$, $f_{SN}^{(\alpha)}(x)$ and $f_{SN}^{(\alpha_o)}(x)$ are k-equivalent. There is a set X_α, which has probability zero, such that if x, y are not in X_α, $\ell_\alpha(x) \geq \ell_\alpha(y)$ if and only if $\ell_{\alpha_o}(x) \geq \ell_{\alpha_o}(y)$. The union X of all X_α with rational α also has probability zero, since the rational numbers are countable. Furthermore, if x and y are not in X, then $\ell_\alpha(x) \geq \ell_\alpha(y)$ for any rational value of α implies $\ell_{\alpha_o}(x) \geq \ell_{\alpha_o}(y)$, and $\ell_{\alpha_o}(x) \geq \ell_{\alpha_o}(y)$ implies $\ell_\alpha(x) \geq \ell_\alpha(y)$ for all rational values of α. Since $f_{SN}^{(\alpha)}(x)$ is continuous in α, $\ell_\alpha(x)$ must be continuous in α also, and it must follow that for any real α in $[a, b]$ and for any x, y not in X, $\ell_\alpha(x) \geqq \ell_\alpha(y)$ if and only if $\ell_{\alpha_0}(x) \geqq \ell_{\alpha_0}(y)$. Then it is easy to show that if x and y are not in X,

$$\int_a^b \left[\ell_\alpha(x) - \ell_\alpha(y)\right] dF(\alpha) \geqq 0$$

if and only if $\ell_{\alpha_0}(x) \geqq \ell_{\alpha_0}(y)$, and hence $\int_a^b f_{SN}^{(\alpha)}(x)\, dF(\alpha)$ is equivalent to $f_{SN}^{(\alpha_0)}(x)$

BIBLIOGRAPHY

On Statistical Approaches to the Signal Detectability Problem:

1. Lawson, J. L., and Uhlenbeck, G. E., Threshold Signals, McGraw-Hill, New York, 1950.

   This book is certainly the outstanding reference on threshold signals. It presents a great variety of both theoretical and experimental work. Chapter 7 presents a statistical approach of the criterion type for the signal detection problem, and the idea of a criterion which minimizes the probability of an error is introduced. (This is a special case of an optimum criterion of the first type.)

2. Davies, I. L., "On Determining the Presence of Signals in Noise," Proc. I.E.E. (London), Vol. 99, Part III, pp. 45-51. March, 1952.

3. Woodward, P. M., and Davies, I. L., "Information Theory and Inverse Probability in Telecommunication," Proc. I.E.E. (London), Vol. 99, Part III, p. 37, March, 1952.

4. Woodward, P. M., and Davies, I. L., "A Theory of Radar Information," Phil. Mag., Vol. 41, p. 1001, 1950.

5. Woodward, P. M., "Information Theory and the Design of Radar Receivers," Proc. I.R.E., Vol. 39, p. 1521.

   Woodward and Davies have introduced the idea of a receiver having a posteriori probability as its output, and they point out that such a receiver gives a maximum amount of information. They have handled the case of an arbitrary signal function known exactly or known except for phase with no more difficulty than other authors have had with a sine wave signal. Their methods serve as a basis for the second part of this report.

6. Reich, E., and Swerling, P., "The Detection of a Sine Wave in Gaussian Noise," Jour. App. Phys., Vol. 24, p. 289, March, 1953.

   This paper considers the problem of finding an optimum criterion of the second type presented in this report for the case of a sine wave of limited duration, known amplitude and frequency, but unknown phase in the presence of Gaussian noise of arbitrary autocorrelation. The method probably could be extended to more general problems. On the other hand, the methods of this report can be applied if the signals are band limited even in the case of non-uniform noise by putting the signals and noise through an imaginary filter to make the noise uniform before applying the theory. See The Theory of Signal Detectability, Part II, Section 3.

7. Middleton, D., "Statistical Criteria for the Detection of Pulsed Carriers in Noise, Jour. Appl. Phys., Vol. 24, p. 371, April, 1953.

A thorough discussion is given of the problem of detecting pulses (of unknown phase) in Gaussian noise. Both types of optimum criteria are discussed, but not in their full generality. The sequential type of test is discussed also.

8. Slattery, T. G., "The Detection of a Sine Wave in Noise by the Use of a Non-Linear Filter," Proc. I.R.E., Vol. 40, p. 1232, October, 1952.

This article considers the problem of detecting a sine wave of known duration, amplitude, and frequency, but unknown phase in uniform Gaussian noise. The article contains several errors, and although the results appear to be correct, they are not clearly presented.

9. Hanse, H., "The Optimization and Analysis of Systems for the Detection of Pulsed Signals in Random Noise," Doctoral Dissertation (MIT), January, 1951.

10. Schwartz, M., "A Statistical Approach to the Automatic Search Problem," Doctoral Dissertation (Harvard), June, 1951.

These dissertations both consider the problem of finding the optimum receiver of the criterion type for radar type signals.

11. North, D. O., "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems," RCA Laboratory Report PTR-6C, 1943.

The ideas of false alarm probability and probability of detection are introduced. North argues that these probabilities will be most favorable when peak signal to average noise ratio is largest. The ideal filter, which maximizes this ratio, is derived. (This commentary is based on second-hand knowledge of the report.)

12. Kaplan, S. M., and Fall, R. W., "The Statistical Properties of Noise Applied to Radar Range Performance," Proc. I.R.E., Vol. 39, p. 56, January, 1951.

The ideas of false alarm probability and probability of detection are introduced and an example of their application to a radar receiver is given.

On Statistics:

13. Neyman, J., and Pearson, E. S., "On the Problem of the Most Efficient Tests of Statistical Hypotheses," Phil. Trans. Roy. Soc., Vol. 231, Series A, p. 289, 1933.

14. Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, 1951.

On Related Topics:

15. Dwork, B. M., "Detection of a Pulse Superimposed on Fluctuation Noise," Proc. I.R.E., Vol. 38, p. 771, July, 1950.

16. Harrington, J. V., and Rogers, T. F., "Signal-to-Noise Improvement Through Integration in a Storage Tube," Proc. I.R.E., Vol. 38, p. 1197, October, 1950.

17. Harting, A. E., and Meade, J. E., "A Device for Computing Correlation Functions," Rev. Sci. Inst., Vol. 23, 347, 1952.

18. Lee, Y. W., Cheatham, T. P., Jr., and Wiesner, J. B., "Applications of Correlation Analysis to the Detection of Periodic Signals in Noise," Proc. I.R.E., Vol. 38, p. 1165, October, 1950.

19. Levin, M. J., and Reintzes, J. F., "A Five Channel Electronic Analog Correlator," Proc. Nat. El. Conf., Vol. 8, 1952.

20. Rice, S. O., "Mathematical Analysis of Random Noise," B.S.T.J., Vol. 23, p. 282-332 and Vol. 24, p. 46-156, 1945-6.

21. Shannon, C. E., "Communication in the Presence of Noise," Proc. I.R.E., Vol. 37, pp. 10-21, January, 1949.

LIST OF SYMBOLS

A	The event "The operator says there is signal plus noise present," or a criterion, i.e., the set of receiver inputs for which the operator says there is a signal present.
$A_1(\beta)$	Any criterion A which maximizes $P_{SN}(A) - \beta P_N(A)$, i.e., an optimum criterion of the first type.
$A_2(k)$	Any criterion A for which $P_N(A) \leq k$, and $P_{SN}(A)$ is maximum, i.e., an optimum criterion of the second type.
CA	The event "The operator says there is noise alone."
d	A parameter describing the ability of a receiver to detect signals. (See Section 5.1 and Fig. 5.1.)
E, $E(s)$	The signal energy.
E^n	The n-dimensional Euclidean space.
$f_N(x)$	The probability density for points x in R if there is noise alone.
$f_{SN}(x)$	The probability density for points x in R if there is signal plus noise.
$F_N(\beta)$, $F_N(\ell)$	The complementary distribution function for likelihood ratio if there is noise alone, i.e., $F_N(\beta)$ is the probability that the likelihood ratio will be greater than β if there is noise alone.
$F_{SN}(\beta)$, $F_{SN}(\ell)$	The complementary distribution function for likelihood ratio if there is signal plus noise.
k	A symbol used primarily for the upper bound placed on false alarm probability $P_N(A)$ in the definition of the second kind of optimum criterion.
$\ell(x)$	The likelihood ratio for the receiver input x. $\ell(x) = \dfrac{f_{SN}(x)}{f_N(x)}$.
n	The dimension of the space of receiver inputs. $n = 2WT$.
N	The event "There is noise alone," or the noise power.
N_o	The noise power per unit bandwidth. $N_o = N/W$.
$P_N(A)$	The probability that the operator will say there is signal plus noise if there is noise alone, i.e., the false alarm probability.

$P_{SN}(A)$	The probability that the operator will say there is signal plus noise if there is signal plus noise, i.e., the probability of detection.
$P(SN)$	The a posteriori probability that there is signal plus noise present. (See Sections 1.3 and 2.3.)
$P_S(s)$	The probability measure defined on R for the set of expected signals.
R	The space of all receiver inputs. (The set of all possible signals is the same space.)
s	A signal s(t), which may also be considered as a point s in R with coordinates $(s_1, s_2, \ldots, s_n)$.
SN	The event "There is signal plus noise."
t	Time.
T	The duration of the observation.
W	The bandwidth of the receiver inputs.
x	A receiver input x(t), which may also be considered as a point x in R with coordinates $(x_1, x_2, \ldots, x_n)$
β	A symbol usually used for the likelihood ratio level of an optimum criterion.
$\mu_{SN}(z)$	The mean of the random variable z if there is signal plus noise.
$\mu_N(z)$	The mean of the random variable z if there is noise alone.
$\sigma_N^2(z)$	The variance of the random variable z if there is noise alone.
σ_N^2	The variance of likelihood ratio if there is noise alone.